

Self-supervised Low Light Image Enhancement and Denoising

Yu Zhang Xiaoguang Di Bin Zhang Qingyan Li Shiyu Yan
Chunhui Wang
Department of Electronic Science and Technology, Harbin Institute of Technology
zhangyuhit2@hit.edu.cn

Abstract

This paper proposes a self-supervised low light image enhancement method based on deep learning, which can improve the image contrast and reduce noise at the same time to avoid the blur caused by pre-/post-denoising. The method contains two deep sub-networks, an Image Contrast Enhancement Network (ICE-Net) and a Re-Enhancement and Denoising Network (RED-Net). The ICE-Net takes the low light image as input and produces a contrast enhanced image. The RED-Net takes the result of ICE-Net and the low light image as input, and can re-enhance the low light image and denoise at the same time. Both of the networks can be trained with low light images only, which is achieved by a Maximum Entropy based Retinex (ME-Retinex) model and an assumption that noises are independently distributed. In the ME-Retinex model, a new constraint on the reflectance image is introduced that the maximum channel of the reflectance image conforms to the maximum channel of the low light image and its entropy should be the largest, which converts the decomposition of reflectance and illumination in Retinex model to a non-ill-conditioned problem and allows the ICE-Net to be trained with a self-supervised way. The loss functions of RED-Net are carefully formulated to separate the noises and details during training, and they are based on the idea that, if noises are independently distributed, after the processing of smoothing filters (e.g. mean filter), the gradient of the noise part should be smaller than the gradient of the detail part. It can be proved qualitatively and quantitatively through experiments that the proposed method is efficient.

1. Introduction

Images captured in low light conditions always suffer from low contrast, low brightness and serious noise, and so on. Low light image enhancement method is used to solve those problems before high-level computer vision tasks, but there are few methods to deal with these problems well at the same time. Recently, various deep learning based al-



Figure 1. Visual comparison with supervised low light image enhancement method KinD [42]. The proposed method can well improve the contrast and brightness of the images and at same time reduce noise, and sharpen the edges. (Best viewed on high-resolution displays with zoom-in)

gorithms have achieved surprising results in some image processing and computer vision tasks, such as object detection [24], [33], [32], [12], image segmentation [12], [25], [2], etc. One most important reason for the rapid development of deep learning in these tasks is that we can obtain a large number of data sets with clear and unambiguous labels. In these tasks, although the construction of the data set requires some cost, it is still acceptable, also on the Internet, a large number of open-source data sets can be found for these tasks to support the training of the network.

However, in low-level image processing tasks such as low light image enhancement, image dehazing, and image restoration, etc., it is difficult to obtain a large number of true input/label image pairs. As for low light image enhancement task, in the previous work, some supervised solutions such as synthesizing low light images [26], using images with different exposure time [4], and so on, have achieved good visual effects, especially in noise reducing. Even though, there are still two problems with those meth-

ods. One is how to ensure that the pre-trained network can be used for images collected from different devices, different scenes, and different lighting conditions rather than building new training data set(*e.g.* [42] failed to remove noises in the background in Fig. 1). The other is how to determine whether the normal light image used for supervision is the best, there can be lots of normal light images for one low light image. Usually, the data sets with paired low/high light images are built with artificial adjustment, which will cost lots of time and energy, and also we cannot make sure the normal light images can complete the training task very well.

In this paper, to overcome the problems in previous works, we proposed a self-supervised low light image enhancement framework to realize image contrast enhancement and denoising at the same time. Similar to previous works[26], two networks are used to achieve contrast enhancement and denoising, respectively. However, different from the previous work, the two networks, *i.e.* Image Contrast Enhancement Network (ICE-Net) and the Re-Enhancement and Denoising Network (RED-Net), are trained self-supervised. And the RED-Net is designed to reduce noises through re-enhancing the contrast of low light images, which can reduce the loss of information caused by pre-/post-processing with existing denoising methods.

For the training of ICE-Net, a Maximum Entropy based Retinex (ME-Retinex) model was proposed. Different from previous Retinex models which only assume that illumination is smooth, in the ME-Retinex model, we introduce a new constraint on the reflectance image that the maximum channel of the reflectance image conforms to the maximum channel of the low light image and its entropy should be largest. With a constraint on reflectance, we can directly control the image enhancement level and convert the ill-conditioned decomposition of reflectance and illumination into a non-ill-posed problem.

For the training of RED-net, we adopt the assumption that the noises conform to the Poisson distribution and in different pixels they are independent. With this assumption, the gradients of noises should be smaller than the gradients of details after the processing of smoothing filters, and most of them are even close to zero. Then it is possible for us to separate most noises and details in reflectance by treating those gradients calculated from smoothed reflectance as weights. At the same time, considering that our task is to enhance the image contrast, and edges with higher gradients often have higher contrast, so the loss functions of RED-Net can be designed to make the gradients of details and edges higher to achieve better contrast enhancement, and it is different from the previous works which only try to preserve edges. Based on those ideas, some self-supervised loss functions are formulated and their effectiveness are proved through experiments.

The loss function in this paper can complete self-supervised training, which means that we can directly solve image enhancement task for one specific image, whether by CNN(Convolutional Neural Network) or analytical methods. However, more training data in CNN often leads to better results(*e.g.* Fig.4), and most of time CNN spends less processing time than analytical methods. The proposed method is independent of the way acquiring low light images, and the training process is completely self-supervised, so the method proposed in this paper has good generalization ability, even if the pre-trained network is not well enough in a new environment, retraining or fine-tuning it without building paired/unpaired normal light images data set is possible for the network. Our contributions can be summarized as:

- We proposed a framework for enhancing low light images, which can enhance the image contrast and reduce noise at the same time. Through the close coupling of the two, we can reduce the loss of information in image enhancement tasks (*e.g.* blur caused by commonly used post-denoising).
- We proposed an Image Contrast Enhancement Network(ICE-Net) and a Re-Enhancement and Denoising Network(RED-Net), and both of them can be trained by self-supervision, which gets rid of the dependence on paired or unpaired images. Also, the RED-Net proposed in this paper can be combined with other Retinex or HSV based image enhancement methods to achieve re-enhancement and noise suppression, even AHE(Adaptive Histogram Equalization [31]) which produces heavy noises, and this is helpful for many previous studies on contrast enhancement.
- We compare the proposed method with several state-of-the-art methods via some comprehensive experiments. And the results are measured by objective indexes and visual quality. All results consistently proved the effectiveness of the proposed method.

2. Related works

Low light image enhancement. Directly adjusting the contrast of the low light image is probably the most intuitive and easy way to realize image enhancement, such as Histogram Equalization(HE), and other improved methods based on HE [30, 36, 28, 3, 19]. Although those improved methods are proposed to achieve noise suppression, hue preserving, brightness preserving, and so on, there are still many problems in directly adjusting the contrast, such as, over- and under- enhancement, noise amplification, *et al.* Gamma correction is another kind of mapping manner, which is also a frequently used method for low light image

enhancement. Although it can promise a well image brightness, and stretch the contrast in low or high areas, it still can not avoid noise amplification and most of the time, its result highly depends on the Gamma value which is chosen artificially.

Retinex is a widely used model for low light imagence-ment in recently years. According to Retinex theory, an image can be decomposed into reflectance and illumination. The early works SSR [15] and MSR [14] treat the reflectance as the final enhancement result. However, since the decomposition is a ill-posed problem and without enough constraints on the reflectance, the enhanced image often have unreal phenomena such as the over-enhancement and whitening. Also, it is hard for those methods to reduce noise. In recent works, illumination are enhanced after the decomposition and the final enhanced image is obtained by recombining the enhanced illumination and reflectance. However, the enhanced image may still have noise and an extra post-denoising procedure have to be preformed[11, 8], which will produce blur in details. [34] introduced a joint low-light enhancement and denoising method, which can achieve denoising and enhancement simultaneously. [21] further improved the method through considering a noise map compared with the conventional Retinex model. Although those methods are proposed to have promising results, most of them need multiple iterations for decomposition which will cost lots of time. Meanwhile, as there is not any method to automatically manipulate the illumination, the enhanced image may not have a proper contrast and usually need careful parameter tuning.

Recently, the amazing performance of deep learning also inspired some promising works in low light image enhancement, including supervised works [26, 37, 42, 6] and unsupervised works [13, 10]. Most of the early works based on supervised learning train the networks with synthetic data sets, such as [26, 20, 38, 35], etc. Although the data obtained by these works seems to be dark and noisy, they are still different from a natural one. Chen et al. [4] introduce a dataset which contains real raw low light images and corresponding raw high light images for training. As there can be lots of reference images for one input low light image, they introduce an amplification ratio in the network to achieve correspondence between the input and reference. This method can well solve the problem of noise and color distortion, however, the ratio must be chosen by user during test which limit the widely use of this method, and there maybe some over-/under-enhanced areas in the image with only one ratio. [37] introduced a dataset named LOL which contains real paired low and high images. And it introduced the Retinex model into the training process to connect the reflection images of the input and reference, and proposed to denoise on reflectance with BM3D[5]. However, it will still cause blur or remain noisy, and it's hard

to find a balance between the two. [42] added a subnet called restoration-net to achieve denoising on reflectance, and provided an extra brightness ratio to control the illumination. However, during the test, it still need to manually adjust the ratio parameter to obtain better enhancement results. Although these methods use real low light data for training, due to the lack of constraint on the contrast of the enhanced image, it can not avoid the problem of over-enhancement (saturation) or under-enhancement in the enhanced image, even with artificial adjustment of parameters. In the unsupervised works, [13] proposed a GAN-based method which can be trained with unpaired data, but it cannot control the enhancement results. [10] proposed a zero-reference low light image enhancement method, which can be trained without any paired or unpaired data. However, it did not provide any noise removal methods.

Image denoising Many denoising methods have been proposed over the past few decades, including conventional methods [5, 9] and learning based methods [40, 23, 1]. However, those denoising methods are not specially designed for the low light image enhancement task. No matter pre-/post-processing with those method will caused details loss, and the learning based method may even invalid for different kind of noise distribution. [41] proposed an denoising method for low light image enhancement, however, it need to be trained with paired low/high light image data. Recently, [18] proposed a unsupervised denoising method named N2V, which can be trained with the noisy image only, however, in our tests, it will still cause blur even re-trained with the enhanced image.

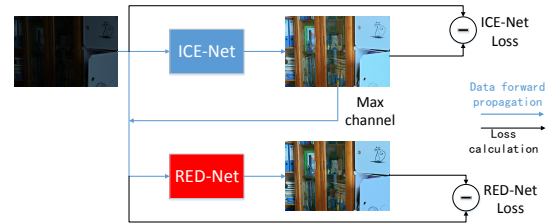


Figure 2. The structure of the proposed method, RED-Net take low light images and the max channel of the output of ICE-Net as input.

3. Proposed Model

The proposed method aims at achieving low light image enhancement and denoising without any artificial adjustment in the case of only having low light images. For example, when the camera enters a new low condition environment, the pre-trained networks may not works for different distribution, and the only data we can get are low light images. In order to achieve automatically contrast enhancement, we proposed a Maximum Entropy based Retinex

model and an self-supervised ICE-Net to take advantage of multiple images. As for the denoising, we proposed an self-supervised RED-Net which is specially designed for the low light image enhancement task. Through the combination of re-enhancement and denoising, the RED-Net can save more details during denoising process. The structure of the proposed method is shown in Fig. 2.

3.1. ME-Retinex model and ICE-Net

Recently, a lot of low-light image enhancement works are based on the following Retinex model:

$$\mathbf{S} = \mathbf{R} \circ \mathbf{I} \quad (1)$$

where \mathbf{S} and \mathbf{I} represent the captured image and the illumination image respectively, \mathbf{R} represents the reflectance and some works treat it as the desired enhanced image, \circ represents the element-wise multiplication. Most of recent works assume that the three color channels of the image have the same illumination in order to simplify the model [37],[42], and the maximum value of the three color channels is generally used as the initial estimate of the illumination map [11]. It has been proved that image enhancement methods based on this simplified Retinex model are equivalent to directly control contrast on V channel in HSV color space and remain the H and S channels unchanged [41]. However, there are still some differences between those two types of methods. Methods which directly control the image contrast usually stretch the contrast of some areas and compress the contrast of other areas, and the compressed areas will lose details(over-/under-enhancement can be treated as details missing). For example, HE will merge the smaller bins, and Gamma correction will compress the contrast of bright areas, both of those will cause the loss of details. And methods based on Retinex usually do not contain the constraints on the contrast of target enhanced image (whether \mathbf{R} or $\mathbf{R} \circ \mathbf{I}^\gamma$), which produce the uncertain of results.

If we transform the HE or Gamma correction to the Retinex model to explain, the enhanced image \mathbf{R} is obtained through \mathbf{S}/\mathbf{I} without the constraint that illumination \mathbf{I} is smooth, then the missing details will be retained in the illumination \mathbf{I} . It can be considered that, if a rich texture area \mathbf{S} is divided by a smooth \mathbf{I} , the details will be in \mathbf{R} which can avoid the loss of details in HE or Gamma correction. Then we can combine the method of directly controlling the contrast with the Retinex model to take advantage of the both. And the combination will suppress the noise through the assumption that illumination is smooth, compared with the method of direct control contrast.

Typically, a Retinex based method can be expressed as follows:

$$\min_{\mathbf{R}, \mathbf{I}} l_{rcon} + \lambda_1 l_{\mathbf{R}} + \lambda_2 l_{\mathbf{I}} \quad (2)$$

Where, l_{rcon} , $l_{\mathbf{R}}$ and $l_{\mathbf{I}}$ represent reconstruction loss, reflectance loss and illumination loss, respectively. λ_1 and λ_2

are weight parameters. The reconstruction loss l_{rcon} can be expressed as:

$$l_{rcon} = \|\mathbf{S} - \mathbf{R} \circ \mathbf{I}\|_1 \quad (3)$$

Where $\|\bullet\|_1$ represents the L_1 norm, we use the L_1 norm to constrain all the losses, and do not compare the impact of L_1 , L_2 , SSIM and other loss functions on low level image processing tasks, since there are already some related studies such as [43].

In this paper, we choose the HE method to form a Maximum Entropy based Retinex model, then the reflectance loss is formulated as:

$$l_{\mathbf{R}} = \left\| \max_{c \in r, g, b} \mathbf{R}^c - F\left(\max_{c \in r, g, b} \mathbf{S}^c\right) \right\|_1 + \lambda \|\nabla \mathbf{R}\|_1 \quad (4)$$

where $F(x)$ means the histogram equalization operator to image x . λ is weight parameters, ∇ means gradient operator. This first term of this loss function means that maximum channel of the reflectance should conform to the maximum channel of the low light image and has the maximum entropy, which can be considered as directly control the contrast of enhanced image. The second term is a commonly used smoothing term to suppress noise, but usually it is hard to distinguish image details and noises well.

For the illumination loss, we adopt the structure-aware smoothness loss proposed in [37]:

$$l_{\mathbf{I}} = \|\nabla \mathbf{I} \circ \exp(-\lambda_3 |\nabla \mathbf{R}|)\|_1 \quad (5)$$

It is proposed that Equation 5 can make the illumination loss aware of the image structure in [37]. And this loss means that the original TV function $\|\nabla \mathbf{I}\|_1$ is weighted with the gradient of reflectance.

For Equations 2-5, we introduce an ICE-Net to solve this optimization problem. Then there will be a question, why introduce CNN to do that? As the ideal image can be obtained from minimizing the total loss, then one could just run this optimization directly on the \mathbf{R} and \mathbf{I} for a single image \mathbf{S} and the introduction of CNN does not seem to be necessary. However, most of the optimization process need multiple iterations which will bring time consumption problems, and with more constraints, the solution will be more complicated. And at the same time, HE is a global enhancement method, which will inevitably lead to the problem of too bright or too dark in some local areas. By introducing CNN and training on multiple images, this problem can be avoided, as shown in Fig. 4. This is because under the HE constraint, the same local area in different images will be enhanced to different degrees, and CNN will be trained to find the median value with L_1 regularization instead of becoming over bright or over dark. In addition, the loss function and ICE-Net is designed to learn how to get an appropriate enhancement in contrast and brightness, so we did not make any special design on denoising and the RED-Net designed in next sub-section can well achieve denoising.

3.2. RED-Net

After the processing of ICE-Net, although the contrast of the image has been improved, there are still some noises in the image. Inspired by [41] which introduce a Conditional Re-Enhancement Network(CRE-Net) to denoise for low light image enhancement tasks, we further propose a self-supervised RED-Net to re-enhance the low light image and denoise at the same time. In this part, we still build the loss function based on Equation 2, however, every sub-loss function has been modified. For the reconstruction loss l'_{rcon} and reflectance loss l'_R in RED-Net, we both adopt the assumption that the noise conforms to the Poisson distribution[39] which is more in line with the real low light image noises. In order to distinguish from the variables in ICE-net, we added the superscript ' for the variables in RED-Net and the reconstruction loss can be expressed as:

$$l'_{rcon} = \mathbf{R}' \circ \mathbf{I}' - \mathbf{S} \circ \log(\mathbf{R}' \circ \mathbf{I}') \quad (6)$$

where \mathbf{R}' and \mathbf{I}' represents the reflectance and illumination produced by RED-Net, respectively, and \mathbf{R}' is also target enhanced image of whole proposed method.

We argue that an image can be divided into different components, including noise, flat area, details and structure information, and there is no clear dividing line between details and structure information. There are many methods to remove noise, but the key to the problem is how to separate details and structure from noise, then to preserve or even strengthen those details and structure during denoising. In order to make reflectance less noise, and preserve rich details and sharp edges, we design the reflectance loss as follows:

$$l'_R = \max_{c \in r, g, b} \mathbf{R}'^c - \max_{c \in r, g, b} \mathbf{R}^c \circ \log(\max_{c \in r, g, b} \mathbf{R}'^c) + \lambda \|\mathbf{W} \circ N(|\nabla \mathbf{R}'|) \circ \exp(-\lambda_3 \mathbf{W} \circ N(|\nabla \mathbf{R}'|))\|_1 \quad (7)$$

where $N(x)$ and $|x|$ represent the local normalization on x and absolute value of x , respectively. \mathbf{R} and \mathbf{R}' represent the output reflectance of the ICE-Net and RED-Net, respectively. \mathbf{W} represents weights, which can be calculated as follows:

$$\mathbf{W} = N(|\nabla(G(\mathbf{R}'))|) \quad (8)$$

where, $G(x)$ represents smooth filter on x (The mean filter are used in proposed method). The graph made by the second term $x * \exp(-\lambda x)$ is shown in Fig. 3. Intuitively, after smoothed, there are still gradients in the details and structure, even they are smaller than before. But the noise and smooth areas may have no gradients or have much smaller gradients. Then we can use the gradients of those smoothed images as the weight. As it can be seen in Fig. 3, when the loss function is in the form of $x * \exp(-\lambda x)$, small x will become smaller, and high x will become higher during

training. And through the local normalization, the details and structure are more likely to fall on the right and make them more sharper during training. As shown in Fig. 5, noise are well removed and the details are preserved.

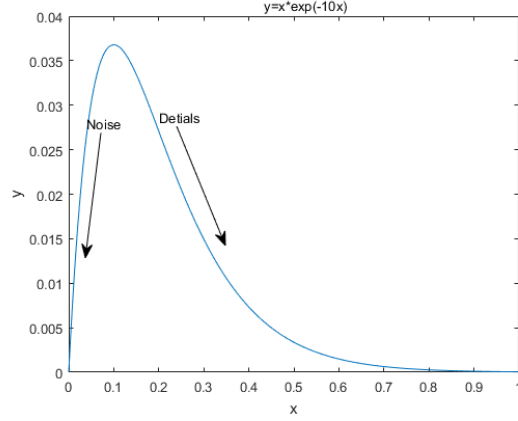


Figure 3. The curve of $y = x * \exp(-\lambda x)$

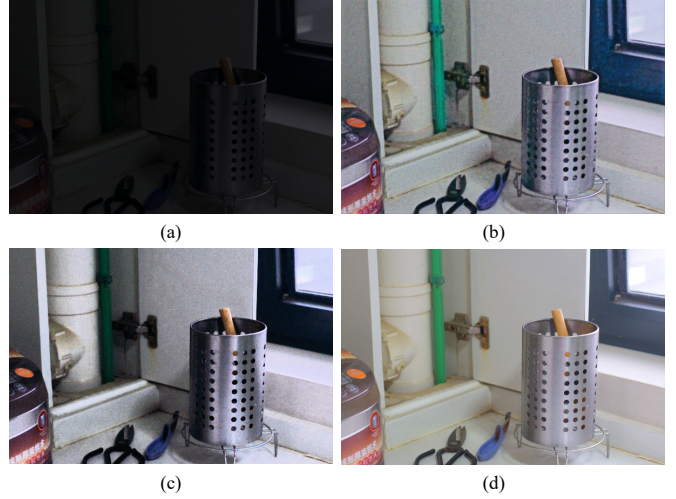


Figure 4. The results by ICE-Net with different training data. (a) Input. (b) The network was trained with multiple data. SSIM:0.6743, PSNR:23.4716, NIQE:3.9140 (c) The network was trained with (a) only. SSIM:0.4858, PSNR:15.5112, NIQE:4.9367 (d) Reference

Typically, illuminations are usually expected to retain only structural information and ignore detailed information. Therefore we can adopt a design similar to reflection loss, and the illumination loss can be expressed as follows:

$$l'_I = \|\mathbf{W}_I \circ N(|\nabla \mathbf{I}'|) \circ \exp(-\lambda_3 \mathbf{W}_I \circ N(|\nabla \mathbf{I}'|)) \circ \exp(-\lambda_3 \mathbf{W}_R \circ N(|\nabla \mathbf{R}'|))\|_1 \quad (9)$$

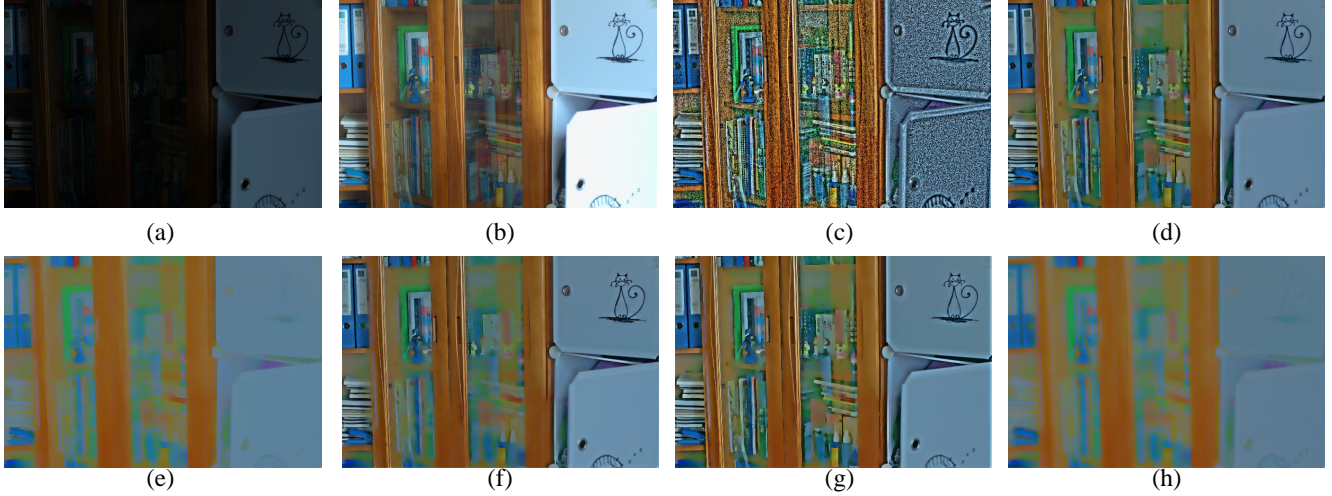


Figure 5. The results generated with different loss in RED-Net. (a) Input low. (b) Reference high. (c) AHE[31]. (d) AHE & RED-Net. (e) w/o \mathbf{W} . (f) w/o \mathbf{W}_I and \mathbf{W}_R . (g) w/o $\exp(-\lambda_3 \mathbf{W}_I \circ N(|\nabla \mathbf{I}'|))$. (h) w/o $\exp(-\lambda_3 \mathbf{W}_R \circ N(|\nabla \mathbf{R}'|))$.

where \mathbf{W}_I and \mathbf{W}_R represents weights, which can be calculated as follows:

$$\mathbf{W}_I = N(|G(\nabla \mathbf{I}')|) \quad (10)$$

$$\mathbf{W}_R = N(|G(\nabla \mathbf{R}')|) \quad (11)$$

where $G(x)$ and $N(x)$ still represent smooth filter and local normalization on x , respectively. Different from \mathbf{W} in Equation 8, the order of gradient operation ∇ and smoothing operation G are switched. It can be considered that, for noise and details, the mean value of the gradient in local area should be small, which is quite different for the structure. For example, texts in a white paper may have opposite gradients in a local area, which makes the mean gradient close to zero. Then with the \mathbf{W}_I and the special design loss form $x * \exp(-\lambda x)$, we can separate the noise and details from the structure in illumination, and make the structure edge sharper during training. Also we preserve $\exp(-\lambda_3 \mathbf{W}_R \circ N(|\nabla \mathbf{R}'|))$ and introduce the weight \mathbf{W}_R to ensure the consistency of the structural information of the reflectance and the illumination. It should be noted that all weight items \mathbf{W} , \mathbf{W}_R , \mathbf{W}_I , do not participate in the back propagation process during training.

4. Experiments

We use the LOL database [37] which contains 500 low/normal light image pairs, 485 of which are used for training and each image size is $400 * 600$. Note that during the training process, we only use natural low light images without any synthetic data and normal light images. During the training process, our batch size is set to 16 and the patch size is set to $48 * 48$. We use Adam stochastic optimization

[17] to train the network and the update rate is set to 0.001. The training and testing of the network are completed on a Nvidia GTX 2080Ti GPU and Inter Core i9-9900K CPU, and the code is based on the tensorflow framework.

To evaluate the performance of the proposed method on enhancing low-light images, we quantitatively and visually compare our method with some low light image enhancement methods, including LIME[11], RRM [21], Retinex-Net[37], KinD [42], and also we collected some data from other data sets for testing.

Three metrics are adopted for quantitative comparison, which are Peak Signal-to-Noise Ratio(PSNR), Structural Similarity(SSIM) [44], and NIQE [27]. NIQE is a non-reference image quality assessment method, which can evaluate the naturalness of the image and a lower value indicates better quality. While, PSNR and SSIM are referenced image quality assessment methods, which indicate the noise level and the structure similarity between the result and the reference, respectively.

4.1. Ablation Study

In this part, to prove the necessity of introducing the CNN and the effectiveness of each component of the proposed method, we have made two ablation studies.

Contribution of ICE-Net. This ablation study is to answer the question that why not just optimize the loss function to get the result, like other variational based Retinex models[16, 29], if the network can be trained in a self-supervised way. As mentioned in Sec. 3.1, the CNN based method ICE-Net is introduced in our proposed method to avoid the problems caused by HE through training with multiple data. Considering it is difficult to directly solve Equation 2 through variational methods under proposed loss functions in this paper, we use a CNN trained with only a

single low light image instead, and the result can be considered as a solution to the Equation 2.

In Fig. 4, we present the results of our ICE-Net trained with one single low light image and with multiple images, and the results show that training with multiple data has a better enhancement effect in contrast and brightness. It can be seen in Fig.4(c), optimization on a single low light image cannot avoid the problems of under-enhancement or over-enhancement (e.g. green pipes and metal hinges), which is caused by HE. However, in Fig. 4(b), training with multiple images, every local area of the enhanced image have a more proper brightness. Also it can be seen through objective indexes, training with multiple data shows better results in PSNR, SSIM and NIQE, which means that the enhanced result with multiple training data has less noise and seems more like reference and is more in line with natural images.

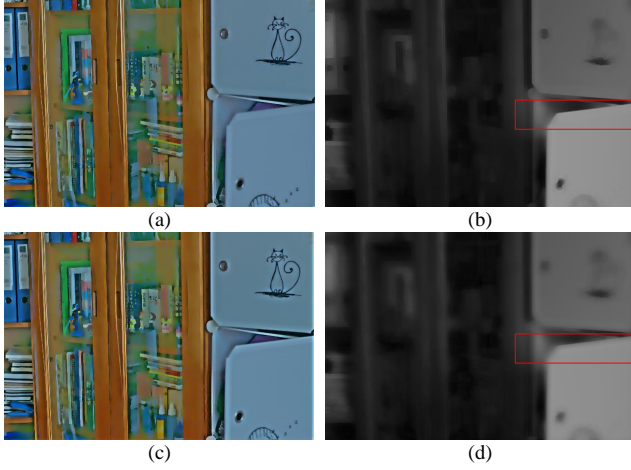


Figure 6. Visual comparison with different loss in RED-Net.(a)-(b) and (c)-(d) are training with or without $\exp(-\lambda_3 \mathbf{W}_I \circ N(|\nabla \mathbf{I}'|))$, respectively. Please zoom in to see the details

Contribution of Each loss in RED-Net. We present the results of RED-Net trained by different losses and weights in Fig. 5. In order to better illustrate the importance of each part of the loss function, we take the AHE[31] which produces serious noise during processing as the contrast enhancement method(e.g. Fig. 5(c)), and study the re-enhancement and denoising effects under different loss functions. We use the complete loss function as the baseline (Completely contains Equation 6 to 11), and study the influence of removing different weight terms, including

- without \mathbf{W} in Equation 7, (Fig. 5 (e))
- without \mathbf{W}_I and \mathbf{W}_R in Equation 9, (Fig. 5 (f))
- without $\exp(-\lambda_3 \mathbf{W}_I \circ N(|\nabla \mathbf{I}'|))$ in Equation 9, (Fig. 5 (g))

- without $\exp(-\lambda_3 \mathbf{W}_R \circ N(|\nabla \mathbf{R}'|))$ in Equation 9, (Fig. 5(h))

As shown in Fig.5 (d), with all the proposed loss functions, the RED-Net can obviously reduce noise and at the same time preserve details. When we simply remove \mathbf{W} (e.g. Fig.5 (e)), only the obvious struture are preserved, that proves the importance and effectiveness of separating the noise and details through \mathbf{W} . And when we remove $\exp(-\lambda_3 \mathbf{W}_R \circ N(|\nabla \mathbf{R}'|))$ (e.g. Fig. 5 (h)), the details are lost and some obvious edges are slightly blurred. And when we remove \mathbf{W}_I and \mathbf{W}_R which are designed to smooth noise and details and preserve structure in illumination, some details in reflectance are blurred(e.g. Fig.5 (f)), which proves the importance of smoothing in the part of noises and details in illumination, and also proves the effectiveness of our design.

In Fig. 5 (g) and (d), it seems that the third kind of loss in this ablation study which removes $\exp(-\lambda_3 \mathbf{W}_I \circ N(|\nabla \mathbf{I}'|))$ in illumination loss does not affect the result of reflectance. However, it can be seen in Fig.6 (b) and (d), the edges (e.g. edge in red rectangles) in illumination are blurred under the third kind of loss, which may caused halo effect in reflectance, and a well illumination can help a lot in future work too(e.g. avoiding over-enhancement). We also study the case that Poisson distribution is not used. However, with AHE [31], the output of the RED-Net is totally unacceptable, and even the structural cannot be saved.

4.2. Comparison with State-of-the-Arts

In this subsection we compare the performance of the proposed method with current state-of-the-art methods through qualitative and quantitative experiments. And during these experiments, we used not only the LOL dataset, but also some standard datasets collected from previous works, including LIME[11](10 images), MF(10 images), and VV¹ (23 images).

We have compared the combination of ICE-Net and RED-Net with some previous methods which can achieve contrast enhancement and denoising, including LIME[11] which has a denoising post-processing, RRM[22] which can jointly enhance contrast and denoise, Retinex-Net[37] which is trained through supervised ways and denoise with BM3D[5] in reflectance, KinD[42] which is trained through supervised ways in contrast enhancement and denoising, and the code is download from the author's homepage and parameters are set as recommended in those paper. The results are shown in Fig. 7 and Table 1 and 2.

Fig.7 shows the qualitative evaluation results, it can be seen that, compared with the LIME and Retinex-net which denoise on the reflectance with BM3D, the two-stage

¹<https://sites.google.com/site/vonikakis/datasets/challenging-dataset-for-enhancement>

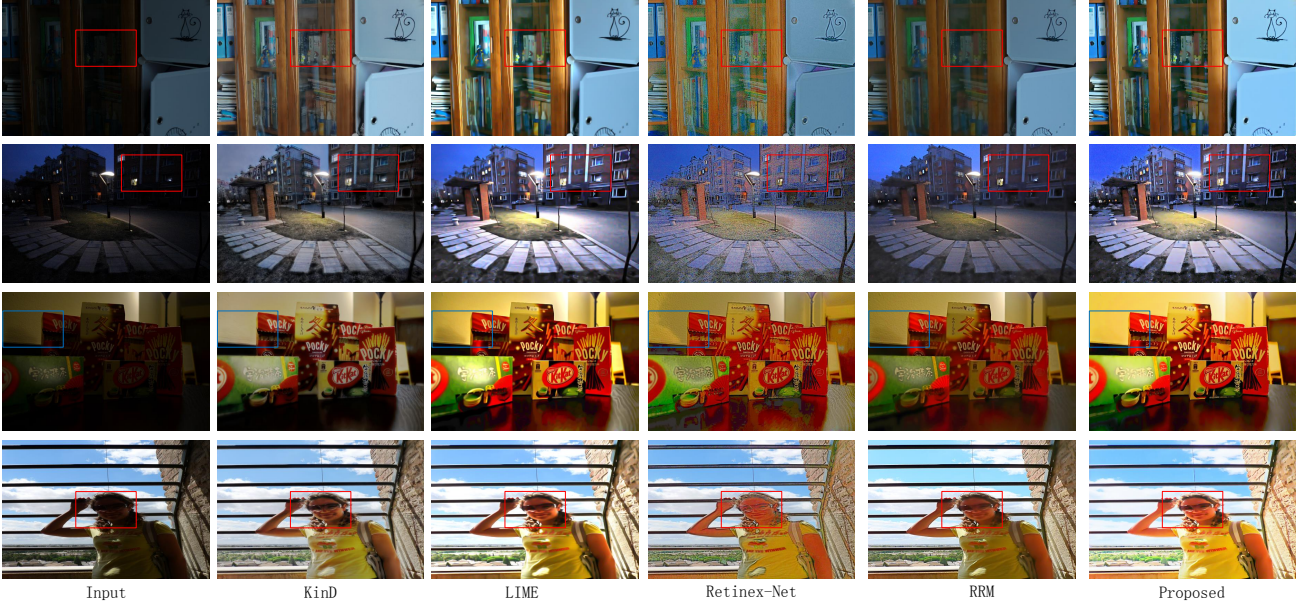


Figure 7. Visual comparison with other state-of-the-art methods, each row comes from a different data set and each column comes from a different method. From left to right are Input, KinD[42], LIME[11], Retinex-Net[37], RRM[22], and proposed method in this paper. Please zoom in to see the details

method proposed in this paper which first enhances, and then re-enhances and denoises can keep a better balance between denoising and detail preserving, and the method is even comparable to the supervised method KinD (e.g. Books in the bookcase processed by LIME are blurred and processed by Retinex-Net still have serious noise, and both our method and KinD can well preserve the texts in the book). Also it can be seen in the second and last row of Fig. 7, our method is able to work under serious noise and non uniform illumination conditions (e.g. the face and arm in the shadow state are well enhanced). At the same time, since we assume that the difference between detail and noise is that noise is distributed independently, which is not always right, the rough wall was smoothed in the blue rectangle in the third row. (More detailed experiments, comparisons, network structure and parameters are included in the supplementary material. In our experiments, the impact of network structure is not significant, and RED-Net has the same network structure as in [41], and the ICE-Net is similar to RED-Net, but has less layers than RED-Net).

Table 1 and 2 show the quantitative evaluation results, it can be seen that, our method gets poor NIQE, and highest PSNR and middle SSIM, which means that after finally enhanced, the image processed by our method seems different from the natural image and reference, and noises in the images are well removed. This is due to that, during designing the ICE-Net and RED-Net, we mainly consider the automatic adaptation ability of the algorithm to the new environment and the removal of noise, and do not take any natural image prior into loss functions, especially in the RED-Net.

Table 1. NIQE scores on the each subset (LOL [37], LIME [11], MF [7], VV), and smaller NIQE indicate more in line with natural images.

Dataset	LIME[11]	RRM[22]	Retinex-Net[37]	KinD[42]	Proposed
LIME [11]	4.08	4.03	4.37	3.59	5.07
LOL [37]	3.95	3.95	9.06	3.89	4.33
MF [7]	3.44	3.68	3.88	3.31	4.59
VV	3.22	3.31	3.57	2.91	4.01

Table 2. SSIM and PSNR scores on the LOL [37] data set, and higher SSIM and PSNR indicate more in line with reference and less noise, respectively.

Dataset	LIME[11]	RRM[22]	Retinex-Net[37]	KinD[42]	Proposed
PSNR	17.22	13.88	16.82	17.64	18.34
SSIM	0.60	0.66	0.57	0.76	0.65

And at the same time, in the process of enhancement, our goal is to enhance the details of each local area, which is quite different from the reference image obtained by adjusting the exposure time.

5. Conclusion and Future Work

In this paper, aiming at automatically enhancing the low light images and denoising, we create a two-stage framework which enhances the image contrast first and then further re-enhances and denoises. And both of the networks in our method can be trained with a self-supervised way, which means that the proposed method can be used in real new unfamiliar environment and new device. The experimental results on various low light data sets show that our method is comparable with many state-of-the-arts methods

on both visual effect and subjective metrics. Our future works will explore how to restore the color degradation, how to combine the RED-Net and the ICE-Net, and how to combine the low light image enhancement and high-level tasks to further improve real-time performance.

References

- [1] F. Agostinelli, M. R. Anderson, and H. Lee. Adaptive multi-column deep neural networks with application to robust image denoising. *Advances in Neural Information Processing Systems*, pages 1493–1501, 2013. 3
- [2] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017. 1
- [3] Turgay Celik and Tardi Tjahjadi. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, 20(12):3431–3441, 2011. 2
- [4] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3291–3300, 2018. 1, 3
- [5] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 3, 7
- [6] Qingxu Fu, Xiaoguang Di, and Yu Zhang. Learning an adaptive model for extreme low-light raw image processing. *arXiv preprint arXiv:2004.10447*, 2020. 3
- [7] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82–96, 2016. 8
- [8] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2782–2790, 2016. 3
- [9] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2862–2869, 2014. 3
- [10] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong. Zero-reference deep curve estimation for low-light image enhancement. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1777–1786, 2020. 3
- [11] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, 26(2):982–993, 2016. 3, 4, 6, 7, 8
- [12] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1
- [13] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlighten: Deep light enhancement without paired supervision. *arXiv preprint arXiv:1906.06972*, 2019. 3
- [14] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, 6(7):965–976, 1997. 3
- [15] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997. 3
- [16] Ron Kimmel, Michael Elad, Doron Shaked, Renato Keshet, and Irwin Sobel. A variational framework for retinex. *International Journal of computer vision*, 52(1):7–23, 2003. 6
- [17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 6
- [18] A. Krull, T. Buchholz, and F. Jug. Noise2void - learning denoising from single noisy images. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2124–2132, 2019. 3
- [19] Chulwoo Lee, Chul Lee, and Chang-Su Kim. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, 22(12):5372–5384, 2013. 2
- [20] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang. Lightnet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, 104:15–22, 2018. 3
- [21] M. Li, J. Liu, W. Yang, X. Sun, and Z. Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 3, 6
- [22] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 7, 8
- [23] Ding Liu, Bihan Wen, Yuchen Fan, Chen Change Loy, and Thomas S. Huang. Non-local recurrent network for image restoration. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS’18*, page 1680–1689, Red Hook, NY, USA, 2018. Curran Associates Inc. 3
- [24] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In *European conference on computer vision*, pages 21–37. Springer, 2016. 1
- [25] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 1
- [26] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2015. 1, 2, 3
- [27] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20(3):209–212, 2012. 6

- [28] Sarif Kumar Naik and CA Murthy. Hue-preserving color image enhancement without gamut problem. *IEEE Transactions on Image Processing*, 12(12):1591–1598, 2003. 2
- [29] Seonhee Park, Soohwan Yu, Byeongho Moon, Seungyong Ko, and Joonki Paik. Low-light image enhancement using variational optimization-based retinex model. *IEEE Transactions on Consumer Electronics*, 63(2):178–184, 2017. 6
- [30] Etta D Pisano, Shuquan Zong, Bradley M Hemminger, Marla DeLuca, R Eugene Johnston, Keith Muller, M Patricia Braeuning, and Stephen M Pizer. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *Journal of Digital imaging*, 11(4):193, 1998. 2
- [31] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987. 2, 6, 7
- [32] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 1
- [33] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*, pages 91–99, 2015. 1
- [34] X. Ren, M. Li, W. Cheng, and J. Liu. Joint enhancement and denoising method via sequential decomposition. In *2018 IEEE International Symposium on Circuits and Systems (IS-CAS)*, pages 1–5, 2018. 3
- [35] Liang Shen, Zihan Yue, Fan Feng, Quan Chen, Shihao Liu, and Jie Ma. Msr-net: Low-light image enhancement using deep convolutional network. *arXiv preprint arXiv:1711.02488*, 2017. 3
- [36] Yu Wang, Qian Chen, and Baeomin Zhang. Image enhancement based on equal area dualistic sub-image histogram equalization method. *IEEE Transactions on Consumer Electronics*, 45(1):68–75, 1999. 2
- [37] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 3, 4, 6, 7, 8
- [38] Jie Yang, Xinwei Jiang, Chunhong Pan, and Cheng-Lin Liu. Enhancement of low light level images with coupled dictionary learning. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, pages 751–756. IEEE, 2016. 3
- [39] Q. Yang, C. Jung, Q. Fu, and H. Song. Low light image denoising based on poisson noise model and weighted tv regularization. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3199–3203, 2018. 5
- [40] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017. 3
- [41] Yu Zhang, Xiaoguang Di, Bin Zhang, Ruihang Ji, and Chunhui Wang. Better than reference in low light image enhancement: Conditional re-enhancement networks. *arXiv preprint arXiv:2008.11434*, 2020. 3, 4, 5, 8
- [42] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 1632–1640, 2019. 1, 2, 3, 4, 6, 7, 8
- [43] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on computational imaging*, 3(1):47–57, 2016. 4
- [44] Zhou Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6